International Journal of Engineering and Techniques - Volume 3 Issue 6, Nov - Dec 2017

RESEARCH ARTICLE OPEN ACCESS

A REVIEW ON BANKING SECTOR

Mr. Subash Babu Bathala[1], Dr. Muthuluru Nagendra[2], Dr.Mahesh Kandakatla[3]

[1]Research Scholar ,Department of Computer Science and Technology,SK University, Anathapur
[2]Professor ,Department of Computer Science and Technology,SK University, Anathapur
[3]Associate Professor,Department of CSE,Vaagdevi College of Engineering, Warangal

Abstract:

Fraud not most effective causes unattainable financial losses but additionally pushes the organization by many steps backwards on this competitive world. Any deliberate act of deceit involving monetary bills or misappropriation of organizations‟ assets for private enrichment is termed as “financial fraud”. In present scenario, enforcing powerful fraud prevention techniques before everything location and detection method in case of failure of preventive measures isn't any more an aggressive gain however a cause that guarantees the survival of the fittest. The probabilities of fraud can be decreased to a degree by way of judging the accuracy of aim and legitimacy of financial transactions, which is almost not possible. For this reason to counter this menace, System must be proactive and consequently put in force fraud prevention and detection strategies. This paper explains the methods to hit upon the fraud in banking sector and how to improve the transaction to discover the frauds.

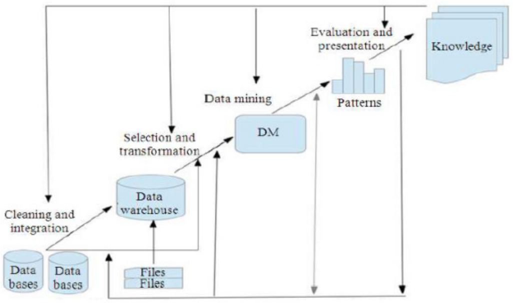
Introduction

Data mining is a process that uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and subsequently gaining knowledge from a large database [1][2]. The knowledge Discovery process, depicts as Fig. 1, consists of following steps which describes how unprocessed data converts into meaningful information [3][4].

- Data Selection. In this step, identifies the location of the data and relevance of the data for the business objectives. Because of electronic data are so pervasive, the quality of data plays a critical role in all business and governmental applications [5][6].
- Data Preparation. Once the data and its location are identified, data cleaning and integration is done in this step. In Data Cleaning, noise data and irrelevant data are removed from the collected data. In Data Integration, different data sources are combined in a common source. Data quality is a major challenge in data mining [7] [8]. Data analysis underlies many computing applications as a part of their on-line operations or in the design phase. Data analysis procedures can be classified as either exploratory or confirmatory, based on the availability of existing and appropriate models for the data source, but the main point of interest in both types of procedures (whether for hypothesis formation or decision-making) is the grouping, clustering or classification of measurements based on either (i) goodness-of-fit to a postulated model, or (ii) natural groupings (clustering) revealed through analysis [9][10].
- Data Transformation. In this step, the selected data is transformed into the form appropriate for the input of data mining process [11][12][13].
- Data Mining. This is the vital step on which effective algorithms [14][15] and techniques applied to process the data into potentially useful patterns to achieve business objectives[16].
- Evaluation. Patterns [17] representing knowledge are identified based on given measures[18][19]
- Representation. The step on which

discovered knowledge visually represents. Visualization techniques[20] are more effective in understanding the output for end users

Data Mining encompasses many different techniques and algorithms. These differ in the kinds of data that can be analyzed and the kinds of knowledge representation used to convey the discovered knowledge [21].



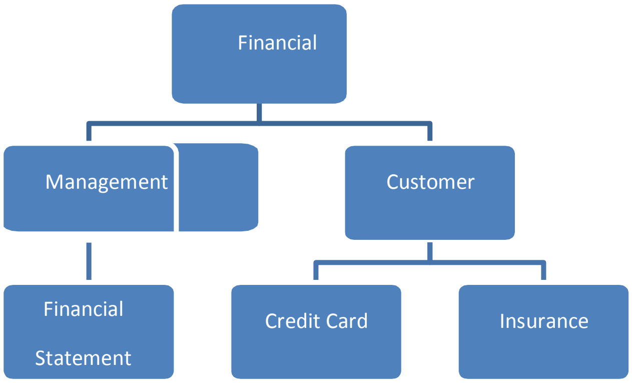
Data mining as a step in the process of knowledge discovery

Data mining being a process of extracting knowledge by learning patterns from the available data has been used widely for developing fraud detection systems. Selection of task relevant data is one of the most important primitive for data mining tasks. Selecting task relevant attribute from large datasets is one of the major hurdles faced by researchers in order to design automated fraud detection systems. It can identify useful and interesting patterns with efficacy, which can be used to find out any inconsistent behavior or fraudulent activity. Researchers from both industry and academia have designed a number of automated/semi - automated data mining systems for detection of financial frauds.

Indian Banking System and Financial Frauds

The Indian banking sector has experienced considerable growth and changes since liberalization of economy in 1991. Though the banking industry is generally well regulated and supervised, the sector suffers from its own set of challenges when it comes to ethical practices, financial distress and corporate governance. This study endeavours to cover issues such as banking frauds and mounting credit card debt, with a detailed analysis using secondary data (literature review and case approach) as well as an interview-based approach, spanning across all players involved in reporting financial misconduct.

In recent years, instances of financial fraud have regularly been reported in India. Although banking frauds in India have often been treated as cost of doing business, post liberalisation the frequency, complexity and cost of banking frauds have increased manifold resulting in a very serious cause of concern for regulators, such as the Reserve Bank of India (RBI). Lokare reveals that the share of retail loan segment in total NPAs continues to stay high, of which credit card loans (2.2 percent) have the third-highest contribution after personal and housing loans[22].



Fraud may be defined as, any intentional act in order to deceive or mislead another person or organization for financial benefits. This deliberate, illegal fraudulent activity may be defined and classified in number of ways depending on type of perpetrators. Fraud committed by individuals external to the organization is termed as customer fraud or external fraud whereas, fraud committed by top - level management is known as management fraud or internal fraud (Fig 2). In this paper, we had classified fraud into two categories namely management fraud and customer fraud.

Management Fraud

An intentional act committed by employees, internal auditors, executives, the board of directors, and managers, who may suffer a financial loss and or reputation loss, is termed as management fraud. In management fraud, CEO"s and executive managements are the perpetrator since they are capable of falsification of expenses, invoices, sales figure etc. Management fraud often called financial statement fraud is a deliberate misstatement of material facts by the management in the books of accounts of a company with the aim of deceiving investors and creditors. This illegitimate task performed by management has a severe impact on the economy throughout the world because it significantly dampens the confidence of investors [23].

Data mining methods are the most widely used technique for detection of financial statement fraud because data mining is capable of extracting novel patterns from large databases by building models, which can further be used for making crucial business decisions. Researchers had applied and evaluated a number of data mining techniques for differentiating fraud and non – fraud organizations. Data set for detection of fraudulent financial reporting consists of financial ratios from publicly available financial results of the organization. Perols Johan L [24] compared the performance of six popular statistical and machine learning models in detecting financial statement fraud under different assumptions of misclassification costs and ratios of fraud firms to non - fraud firms. The fraudulent observations were located based on firms investigated by the SEC for financial statement fraud and reported in Accounting and Auditing Enforcement Releases (AAER) from the fourth quarter of 1998 through the fourth quarter of 2005. A total of 42 predictors were examined, only six were consistently selected and used by different classification algorithms. Ravisankar et al [25] predicted the occurrence of financial fraud by using six data mining techniques by analyzing data from 202 Chinese companies. Thirty-five financial ratios were considered for detecting fraudulent financial reporting.

Gupta et al [26] detected fraudulent financial reporting by using three data mining techniques by analyzing data from 114 listed companies. 63 financial ratios were considered for fraud detection.

Customer Fraud

Acquisition of goods/services resorting to unethical means or deceiving an organization by the customer for personal gains can be termed as customer fraud. In this type of fraud, a customer acquires the goods/services by unethical means or deceives an organization with an intention to commit financial loss. A customer can mislead various financial institution and insurance companies that will result two sub categories of customer fraud namely credit card fraud and insurance fraud.

Credit card fraud

There are many review papers describing the different types of frauds and different fraud detection techniques. Revolution from traditional commerce to ecommerce has compelled the use of credit card on a large scale. According to RBI [27], more than 6 crore of transactions worth Rs. 190989.13 Million went through in May 2015. Unfortunately, this intensifying usage also invites criminals to fraudulently use credit cards to earn money / acquire product or service by unethical means. According to the Nilson Report [28], fraud losses on credit cards, debit cards, and prepaid cards worldwide hit $16.31 billion in 2014 on a total card sales volume of $28.844 trillion. A study released in 2016 by New LexisNexis Risk Solutions [29] revealed that credit card fraud costed $7.6 billion. This rising number is an alarming call to provide some automatic intelligent system that can detect fraud before it is being committed. John [30] states that there is a fixed pattern to how credit-card owners consume their credit card on the internet. From this statement, it can be deduced that if a customer deviates from his normal course of behavior then there is something suspicious. Although not every asymmetrical action ensures fraud but chances are quite high of being deceptive. Keeping a regular watch on all activities of the user can prove to be beneficial in detecting a fraudulent act. The goal of a reliable detection system is to learn the behavior of users dynamically to minimize its own loss. Researchers have used number of attributes from the database, which defines the profile of a customer and pattern of his transactions. John

[30] developed a neural network model, trained with attributes like merchants‟ websites, regular good and services purchased in past transactions of credit card holder; shipping address, email address and phone number of customer and geolocation of transaction. Avinash et al [31] takes into consideration, the factors revealing the cardholder’s spending behavior, i.e. columns related to his past transactions. Jyotindra et al‟s [32] model has taken 10 parameters like category of the purchase, same product purchased within short time, Late night transaction, Overseas transaction etc. from a dataset of Online shopping firm’s credit card transaction data. CardWatch [33] is a neural network based credit card fraud detection which trains a neural network with the past data of particular customer spending behavior and the authors tested their software on synthetically generated data. Ghosh and Reilly [34] developed a multilayer feedforward neural network based fraud detection model for Mellon Bank. Self-organizing map neural network is a clustering technique used by many researchers to detect credit card fraud based on customer behaviour. Duman andOzcelik [35] suggested a combination of genetic algorithm and scatter search for improving the credit card fraud detection and the experimental results show a performance improvement of 200%. Maes et al. [36] have performed experiments using Bayesian belief networks and artificial neural networks on credit card fraud detection. They found that Bayesian network has better fraud detection capability than artificial neural network. Srivastava et al. [37] proposed a credit card fraud detection model using hidden Markov model. In their research they trained the HMM with the normal behaviour of the customer and the incoming transaction is considered as fraudulent if it is not accepted by the HMM with high probability. Chan et al. [38] addressed the issues of skewed distribution of credit card data and nonuniform cost by using a distributed data mining model. Seyedhossein and Hashemi [39] proposed a fraud detection method which extracts the inherent pattern from a credit card time series and uses this pattern for earlier fraud detection. S´anchez et al. [40] suggested and demonstrated the use of fuzzy association rule mining in extracting knowledge useful for fraud detection from transactional credit card databases. Syeda et al. [41] suggested a parallel processing environment for fast training of neural network for fraud detection. Wong et al. [42] investigated the application of artificial immune system in credit card fraud detection even though it was a preliminary research. Lu and Ju [43] designed a class weighted support vector machine for classifying the imbalanced credit card transaction data. Panigrahi et al. [44] integrated three different approaches—rule-based filtering, Dempster-Shafer theory, and Bayesian learning—for reducing false alarms in credit card fraud detection. Jha et al. [45] employed transaction aggregation strategy to capture costumer buying behaviour prior to each transaction and used these aggregations to detect credit card fraud. Most of the work found in the literature works on customer spending behaviour analysis and some of them use some derived attributes also. However, we could not find any research performed on anonymous credit card transaction dataset where the derived attribute concept fails. Thus, the objective of this research was to develop a credit card fraud detection model which can effectively detect frauds from imbalanced and anonymous dataset. In order to handle anonymous data, which is the nature of data generally banks provide due to security reasons, the proposed fraud detection model considered each attribute equally without giving preference to any attribute in the dataset. Also the proposed fraud detection model creates separate legal transaction pattern (costumer buying behaviour pattern) and fraud transaction pattern (fraudster behaviour pattern) for each customer and thus converted the imbalanced credit card transaction dataset into a balanced one to solve the problem of imbalance.

It is evident from the above discussion, that every past literature work has different parameters‟ name depending on the available transactions database. Therefore, while selecting an attribute in consideration for future research work, focus should be on the type of information it is providing i.e. category in which it is being classified rather than restricting to the name of attributes in the database. Various factors one should consider while selecting the effective attributes are broadly classified as under:

Insurance fraud

An act performed by the insured person / beneficiary to apply for compensation by producing fake documents / reports is termed as insurance fraud. According to India forensic Research, every single insurance company loses 8.5% of its revenues to the frauds [46]. Teris Roberts [47]has suggested an suspicious activity assessment for insurance frauds wherein system takes care of risk factors like claim profile, policy profile, customer profile, entity profile and network profile and grades score to all the related entities(customer, broker etc.) at regular intervals.

Automobile insurance fraud includes staged automobile accident and a real accident with fabricated bills, thespian accidents, excessive repairs, and fictitious personal injuries all with one intention in mind i.e. false insurance claims resulting in financial loss to the companies [48]. Rekha [49] collected data from attributes like Policy Holder, Driver Rating, Vehicle Age, Price, and Report Filed. Liu et al [50] used a dataset of 5000 records with six attributes namely age, gender, claim amount, tickets, claim times, and accompanied with attorney. Lovera et al [51] suggests that staged accidents have several common characteristics. They occur in late hours and non- urban areas in order to reduce the probability of witnesses. Drivers are usually younger males; there are many passengers in the vehicles, but never children or elders. The police is always called to the scene to make the subsequent acquisition of means easier.

Thus, it can be concluded that attributes specifying the age and gender of driver, driver rating, age of passengers, and number of prior claims, price, and age of vehicle should be considered while selecting the dataset for further research work.

Conclusion:

Data Mining is a tool and techniques used to extract meaningful information from the collected data, enables financial institutions to make better decision-making process. Based on the standard or rules set by the organization and regulatory authorities, data mining tool extract the knowledge based on the rule set and throws the output in visual tools, thereby making end user life easy to make decisions properly. Banks and Financial organizations started allocating funds and time for implementing data mining tools in the area of decision making by realizing the necessity of data mining in their system. This paper presents a comprehensive analysis of data mining techniques used for detection of each type of financial fraud. This will lay foundation to provide scope for further research in the field of credit card fraud detection.

References:

1. Urban, E., Aronson, J.E., Liang, T.P., &Sharda, R. (2007)." Decision Support and Business Intelligence Systems", Eighth edition, Pearson Education, 2007.
2. Relative Parameter Quantification in Data Mining - A Case Study on Telecom Cellular Mobile Service Providers in Terms of QOS in India by Mahesh Kandakatla, Lokanatha C Reddy,prashanth bolukonda in International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.5, September 2015.
3. Bhambri, V., 2011. Application of data mining in banking sector. IJCST, 2: 199-202.
4. Chen, I., L. Chi-Jie, L. Tian-Shyug and L. Chung-Ta, 2009. Behavioral scoring model for bank customers using data envelopment analysis. Stud. Comput. Intell., 214: 99-104. DOI: 10.1007/978-3-540-92814-0_16
5. Batini, C., C. Cappiello, C. Francalanci and A. Maurino, 2009. Methodologies for data quality assessment and improvement. ACM Comput. Surveys. DOI: 10.1145/1541880.1541883
6. Yap, B.W., S.H. Ong and N.H.M. Hussain, 2011. Using data mining to improve assessment of credit worthiness

via credit scoring models. Expert Syst. Appli., 38: 13274-13283. DOI: 10.1016/j.eswa.2011.04.147

7. Blake, R. and P. Mangiameli, 2011. The effects and interactions of data quality and problem complexity on classification. J. Data Inform. Q. DOI: 10.1145/1891879.1891881

8. Petry, F.E. and L. Zhao, 2009. Data mining by attribute generalization with fuzzy hierarchies in fuzzy databases. Fuzzy Sets Syst., 160: 2206-2223. DOI: 10.1016/j.fss.2009.02.014

9. Shinde, P., 2012. Data mining using artificial neural network tree. IOSR J. Eng.

10. Fung, B.C.M., K. Wang, R. Chen and P.S. Yu, 2010. Privacy-preserving data publishing: A survey of recent developments. ACM Comput. Surveys. DOI: 10.1145/1749603.1749605.

11. Han, J., M. Kamber and J. Pie, 2011. Data Mining Concepts and Techniques. 3rd Edn., Elsevier, Burlington, ISBN-10: 9780123814807, pp: 744.

12. Moin, K.I. and Q.B. Ahmed, 2012. Use of data mining in banking. Int. J. Eng. Res. Applic., 2: 738-742.

13. Prakash, B.V.A., D.V. Ashoka and V.N.M. Aradhya, 2012. Application of data mining techniques for software reuses process. Proc. Technol., 4: 384-389. DOI: 10.1016/j.protcy.2012.05.059

14. Bhambri, V., 2011. Application of data mining in banking sector. IJCST, 2: 199-202.

15. Hsu, F.M., L.P. Lu and C.M. Lin, 2012. Segmenting customers by transaction data with concept hierarchy. Expert Syst. Applic., 39: 6221-6228. DOI: 10.1016/j.eswa.2011.12.005

16. Hammawa, M.B., 2011. Data mining for banking and finance. Oriental J. Comput. Sci. Technol., 4: 273-280.

17. Tremblay, M.C., K. Dutta and D. Vandermeer, 2010. Using data mining techniques to discover bias patterns in missing data. J. Data Inform. Q., 2: 1-19. DOI: 10.1145/1805286.1805288

18. Kontonasios, K.N., E. Spyropoulou and T.D. Bie, 2012. Knowledge discovery interestingness measures based on unexpectedness. Wiley Interdisciplinary Rev.: Data Min. Knowl. Discovery, 2: 386-399. DOI: 10.1002/widm.1063

19. Wikum, G., G. Kasneci, M. Ramanath and F. Suchanek, 2009. Database and information-retrieval methods for knowledge discovery. Mag. Commun. ACM, 52: 56-64. DOI: 10.1145/1498765.1498784

20. Herawan, T. and M.M. Deris, 2011. A soft set approach for association rules mining. Knowl.-Based Syst., 24: 186-195. DOI: 10.1016/j.knosys.2010.08.005

21. Mabroukeh, N.R. and C.I. Ezeife, 2010. A taxonomy of sequential pattern mining algorithms. ACM Comput. Surveys. DOI: 10.1145/1824795.1824798

22. Lokare, S. M. (2014), "Re-emerging stress in the Asset Quality of Emerging Markets: Macro Financial Linkages", RBI Working Papers.

23. Gupta and Nasib S. Gill (2012), "Prevention and Detection of Financial Statement Fraud – An Implementation of Data Mining Framework", International Journal of Advanced Computer Science and Applications, Volume 3 No. 8, pp. 150 – 156, Published by The Science and Information Organization, U.S.A.

24. Johan Perols, Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms, A Journal of Practice &Theory, 30 (2), 19 (2011), pp. 19-50.

25. P. Ravisankar, V. Ravi, G.Raghava Rao, I., Bose, Detection of Financial Statement Fraud and Feature Selection using Data Mining Techniques, Decision Support Systems, Volume 50 (2011), pp. 491 – 500.

26. Gupta and Nasib S. Gill (2012), “Prevention and Detection of Financial Statement Fraud – An Implementation of Data Mining Framework”, International Journal of Advanced Computer Science and Applications, Volume 3 No. 8, pp. 150 – 156, Published by The Science and Information Organization, U.S.A.

27. ATM & Card Statistics for May 2015 by RBI. Available at https://rbi.org.in/Scripts/ATMView.aspx?atmid=51

28. Report available at https://www.nilsonreport.com/upload/pdf/Global_Card_Fraud_Damages_Reach_16B_-_PYMNTS.com.pdf

29. Report available at https://nilsonreport.com/upload/pdf/Card_Fraud_Costing_Issuers_10.9_Billion_Annually_-_Yahoo_Finance.pdf

30. John Akhilomen.”Data Mining Application for Cyber Credit-card Fraud Detection System”; Journal of Engineering Science and Technology Vol. 6, No. 3 (2011) 311 - 322 . Proceedings of the World Congress on Engineering 2013 Vol III, WCE 2013, July 3 - 5, 2013, London, U.K

31. Avinash Ingole, Dr. R. C. Thool .”Credit Card Fraud Detection Using Hidden Markov Model and Its Performance”. International Journal of Advanced Research in Computer Science and Software Engineering. June 2013.

32. Dr. Jyotindra N. Dharwa Dr. Ashok R. Patel. A Data Mining with Hybrid Approach Based Transaction Risk Score Generation Model (TRSGM) for Fraud Detection of Online Financial Transaction. International Journal of Computer Applications (0975 – 8887) Volume 16– No.1, February 2011.

33. Aleskerov, E., Freisleben, B., and Rao, B., 1997. CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection, Proceedings of IEEE/IAFE: Computational Intelligence for Financial Eng. (1997), pp. 220-226.

34. Ghosh, S., and Reilly, D.L., 1994. Credit Card Fraud Detection with a Neural-Network, 27th Hawaii International l Conference on Information Systems, vol. 3 (2003), pp. 621- 630

35. E. Duman and M. H. Ozcelik, “Detecting credit card fraud by genetic algorithm and scatter search,” Expert Systems with Applications, vol. 38, no. 10, pp. 13057–13063, 2011.

36. S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick, “Credit card fraud detection using Bayesian and neural networks,” in Proceedings of the 1st International NAISO Congress on Neuro Fuzzy Technologies, pp. 261–270, 1993.

37. A. Srivastava, A. Kundu, S. Sural, and A. K. Majumdar, “Credit card fraud detection using hidden Markov model,” IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2008.

38. P. K. Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo, “Distributed data mining in credit card fraud detection,” IEEE Intelligent Systems and Their Applications, vol. 14, no. 6, pp. 67–74, 1999

39. L. Seyedhossein and M. R. Hashemi, “Mining information from credit card time series for timelier fraud detection,” in Proceeding of the 5th International Symposium on Telecommunications (IST '10), pp. 619–624, Tehran, Iran, December 2010.

40. D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, “Association rules applied to credit card fraud detection,” Expert Systems with Applications, vol. 36, no. 2, pp. 3630–3640, 2009.

41. M. Syeda, Y.-Q. Zhang, and Y. Pan, “Parallel granular neural networks for fast credit card fraud detection,” in Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '02), vol. 1, pp. 572–577, Honolulu, Hawaii, USA, May 2002

42. N. Wong, P. Ray, G. Stephens, and L. Lewis, “Artificial immune systems for the detection of credit card fraud,” Information Systems, vol. 22, no. 1, pp. 53–76, 2012.

43. Q. Lu and C. Ju, “Research on credit card fraud detection model based on class weighted support vector machine,” Journal of Convergence Information Technology, vol. 6, no. 1, pp. 62–68, 2011.

44. S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, “Credit card fraud detection: a fusion approach using Dempster-Shafer theory and Bayesian learning,” Information Fusion, vol. 10, no. 4, pp. 354–363, 2009.

45. S. Jha, M. Guillen, and J. C. Westland, “Employing transaction aggregation strategy to detect credit card fraud,” Expert Systems with Applications, vol. 39, no. 16, pp. 12650–12657, 2012.

46. India Forensic research available at http://indiaforensic.com/certifications/india-loses-6-25-billion-to-insurance-frauds-an-indiaforensic-research.

47. Terisa, R. "Improving the defense lines: the future of fraud detection in the insurance industry (with fraud risk models, text mining, and social networks)." SAS Global forum, Washington. 2010.

48. E.W.T. Ngai, H. Yong, Y.H. Wong, C. Yijun and S. Xin, “The application of data mining techniques in financial fraud detection: A Classification Framework and an Academic Review of Literature”. Decision Support Systems, Volume 50, Issue 3, February 2011.

49. Bhowmik, Rekha. "Detecting auto insurance fraud by data mining techniques." Journal of Emerging Trends in Computing and Information Sciences 2.4 (2011): 156-162.

50. Jenn-Long Liu and Chien-Liang Chen .”Application of Evolutionary Data Mining Algorithms to Insurance Fraud Prediction”. Proceedings of 2012 4th International Conference on Machine Learning and Computing IPCSIT vol. 25 (2012) © (2012) IACSIT Press, Singapore.

51. Šubelj, Lovro, Štefan Furlan, and Marko Bajec. "An expert system for detecting automobile insurance fraud using social network analysis." Expert Systems with Applications 38.1 (2011): 1039-1052.